HEVC-BASED LIGHT FIELD IMAGE CODING WITH BI-PREDICTED SELF-SIMILARITY COMPENSATION

Caroline Conti, Paulo Nunes, Luís Ducla Soares
Instituto de Telecomunicações; ISCTE - Instituto Universitário de Lisboa, Portugal
e-mails: caroline.conti, paulo.nunes@lx.it.pt, lds@lx.it.pt

ABSTRACT

This paper proposes an efficient light field image coding (LFC) solution based on High Efficiency Video Coding (HEVC). The proposed light field codec makes use of the self-similarity (SS) compensated prediction concept to efficiently explore the inherent correlation of this type of content. To further improve the coding performance, a bi-predicted SS estimation and SS compensation is proposed, where the candidate predictor can be also devised as a linear combination of two blocks within the same search window. In addition, an improved vector prediction scheme is also used to take advantage of the particular characteristics of the SS prediction vectors. Experimental results show that the proposed LFC scheme is able to outperform the benchmark solutions with significant gains.

Index Terms— Light Field, Plenoptic, Holoscopic, Image Coding, HEVC, Self-Similarity, Bi-Prediction

1. INTRODUCTION

Light field imaging based on microlens arrays – also known as holoscopic, plenoptic, and integral imaging – derives from the fundamentals of light field/radiance sampling [1], where not only the spatial information about the three-dimensional (3D) scene is represented but also angular viewing direction, i.e., the "whole observable" scene.

Recently, light field imaging has become a prospective imaging approach for providing richer content capturing, visualization, and manipulation, being applied in many different areas of research, e.g., 3D television [2], [3], image recognition [4], and medical imaging [5]. Among the advantages of employing a light field imaging system is the enabling of new degrees of freedom in terms of content production and manipulation, supporting functionalities not straightforwardly available in conventional imaging systems, namely, post-production refocusing, changing depth-of-field, and changing viewing perspective.

However, introducing light field image and video applications with its appealing functionalities will require providing an efficient coding scheme to deal with the large amount of data involved in such type of systems. In this context, novel initiatives on light field image and video coding standardization are also emerging. Notably, the JPEG committee has recently started the JPEG Pleno standardization initiative [6] which addresses representation and coding of emerging new imaging modalities. In addition, the MPEG group has also started the third phase of Free-viewpoint Television (FTV), targeting free navigation and full parallax imaging applications [7]. Moreover, the grand challenge on light field image compression at the International Conference on Multimedia and Expo (ICME) 2016 [8] evidences the growing relevance of this topic in the research community.

Regarding light field coding (LFC) approaches, although High Efficiency Video Coding (HEVC) Main Still Picture profile [9] presents significant compression performance improvements in comparison to other state-of-the-art still image coding technologies – such as JPEG and JPEG 2000 standards [10], [11], previous work [12], [13] has shown that further improvements are still possible for light field images by integrating a special prediction scheme to exploit the inherent correlations of this type of content. In [12], a coding scheme based on the concept of self-similarity (SS) compensated prediction is proposed to improve the performance of HEVC for light field images. Similar to motion compensation, the SS estimation process uses a block-based matching over the previously coded and reconstructed area of the current picture (referred to as SS reference [12]), to find the 'best' predictor for the current block. As a result, the chosen block becomes the candidate predictor and the relative position between the two blocks is derived as an SS vector. More recently, in [13], a novel vector prediction scheme was also proposed to take advantage of the particular characteristics of the SS prediction data and, then, to increase coding efficiency.

Motivated by the results in [12], [13], this paper proposes an efficient coding solution for light field imaging based on microlens arrays, which is based on HEVC and the SS compensated prediction. To further improve the compression performance compared to the previously proposed solution in [13], a bi-predicted SS estimation and SS compensation is proposed, where the candidate predictor can be devised as a linear combination of two blocks in the same area of the SS reference. The proposed LFC solution using bi-predicted self-similarity, LFC (Bi-SS), and the accompanying material correspond to the authors' response to the ICME 2016 grand challenge on light field image compression [8].

The remainder of this paper is organized as follow: Section 2 presents the LFC (Bi-SS) solution architecture and describes the special coding tools for efficient light field compression; Section 3 presents the experimental results; and, finally, Section 4 concludes the paper.

2. PROPOSED LFC (BI-SS) SOLUTION

Fig. 1 presents the proposed LFC (Bi-SS) solution architecture, which is based on HEVC and comprises additional and modified modules to efficiently handle the light field content. Specifically, the proposed codec introduces an additional type of prediction – the SS prediction – and the encoder will choose the best, among SS and HEVC Intra prediction, based on the conventional rate-distortion optimization (RDO) process.

The proposed LFC (Bi-SS) solution architecture (Fig. 1) aims to explore the particular correlation of the light field content without requiring any explicit knowledge about the used optical system (e.g., microlens' size, focal length, and distance of the microlenses to the image sensor). Notice that, although these parameters may be provided by camera makers, many of them are highly dependent on the manufacturing process, being different from camera to camera (e.g., each microlens may vary slightly in shape, size, and relative position). For this reason, using compression and rendering tools that are less dependent on a very precise calibration process would be advantageous for supporting a wider range of devices without increasing the processing complexity. The proposed solution is also advantageous in terms of the encoder/decoder computational complexity and necessary memory, which are no larger than that for HEVC Inter B frame coding with one active reference in each reference picture list. Detailed information about HEVC computational complexity can be found in [13]. In addition to this, as will be required in JPEG Pleno [6], backward compatibility to JPEG can be easily adapted for the proposed coding architecture by making use of one of the HEVC multi-layer extensions, such as multiview and scalable HEVC, since they guarantee the support for an external base layer (which would be JPEG coded).

The novel and modified blocks of the LFC (Bi-SS) solution (highlighted in Fig. 1), are explained as follows.

2.1. Bi-predicted SS Estimation and Compensation

As a result of the used optical system, the light field raw image corresponds to a 2D array of micro-images (MIs), where both light intensity and direction information are recorded. Due to the small baseline between adjacent microlenses in the acquisition process, a significant cross-correlation exists between neighboring MIs, as shown in Fig. 2a (the constant distance between regular spikes correspond to the MI spacing).

Since this cross-correlation is distributed along the MIs in all directions, it is likely that good predictions will be distributed along these directions. This can be exemplified by the SS vector distribution depicted in Fig. 2b (brighter areas correspond to more frequent SS vector amplitudes), when using the previously proposed SS compensated prediction [13]. As discussed in [14] in the context of temporal bidirectional prediction, by jointly estimating two good predictions, the accuracy of the compensated prediction can be significantly improved compared to the unidirectional prediction[14]. Hence, the performance of the SS compensated prediction is also expected to improve with a bi-predicted approach as it gives a larger number of good possibilities for the encoder to choose in a RDO sense.

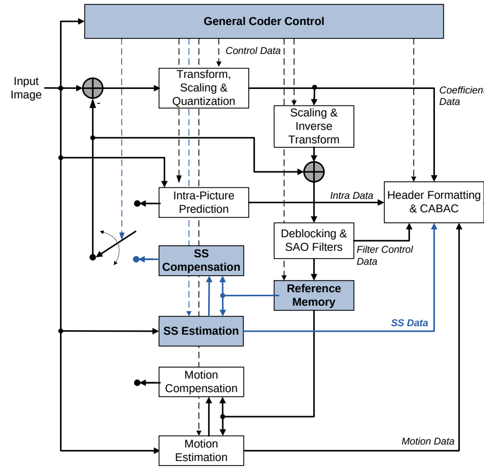

Fig. 1 Coding architecture of the proposed LFC (Bi-SS) solution based on HEVC. The novel blocks are highlighted in blue

Therefore, introducing the SS compensated bi-prediction includes adaptation at the following stages that are managed by the SS estimation and SS compensation blocks in Fig. 1:

1) *Prediction modes:* The proposed coding scheme includes two new prediction modes, referred to as SS and SS-skip prediction modes. These prediction modes can be evaluated for all coding block (CB) sizes (i.e., from 64×64 down to 8×8) in the conventional RDO process to choose the best prediction mode. In the SS-skip prediction mode, the SS vector candidates are derived by using the vector prediction method presented in Section 2.2 and no further information is transmitted.
2) *Partition patterns:* The SS prediction mode allows to use the eight HEVC partition patterns (i.e., 2N×2N, N×2N, 2N×N, N×N, 2N×nU, 2N×nD, nL×2N and nR×2N [9]). Basically, each pattern defines a flexible way to partition the coding block for evaluating the SS estimation process. The SS-skip is employed only for the 2N×2N partition pattern.
3) *SS estimation:* In the SS prediction mode, the causal search window is restricted (see Fig. 2c) to minimize the computational complexity. Two possible candidate predictors are derived to predict the current block and the best among them is chosen in a RDO sense. The first candidate predictor is given by the best block matching to the current block in the search window (the unidirectional predictor). On the other hand, the second candidate predictor is selected by jointly searching for the best linear combination between the first candidate predictor and a second candidate block in the same search window (the bidirectional predictor), as seen in Fig. 2c.
4) *SS compensation:* When using bi-predicted SS compensation, the candidate predictor block is simply derived as the average between the two selected predictor blocks, as exemplified in Fig. 2c.

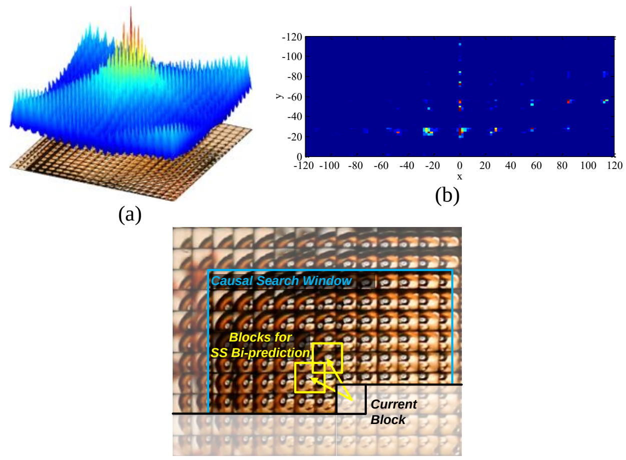

Fig. 2 The SS prediction concept: (a) the inherent MI cross-correlation in a light field image neighborhood; (b) SS vectors distribution; and (c) the SS estimation process with Bi-prediction

2.2. SS Vectors Prediction

To further improve the performance of the conventional HEVC vector prediction methods, a set of new candidate SS vectors, referred to as MI-based SS candidate vectors in [13], is also included into HEVC Advanced Motion Vector Prediction (AVMP) and merge methods [9] so as to force the candidate vectors to be distributed according to the structure of MIs, as well as to be inside the area of previous coded and reconstructed CBs (a detailed description of the MI-based SS candidate vectors selection is given in [13]).

2.3. Reference Picture List Construction

To allow the SS estimation and SS compensation in Intra-coded frames of HEVC, the reference list construction and signaling need to be allowed so as to include the SS reference. This process is similar to the temporal list construction on HEVC Inter-coded frames, and is managed by the reference memory block in Fig. 1.

3. EXPERIMENTAL RESULTS

To evaluate the performance of the proposed LFC (Bi-SS) solution, the common test conditions defined in [8] are adopted, and can be summarized as follow:

- **Test Images:** The twelve LF test images provided by the ICME 2016 grand challenge in light field image compression are used [8].
- **Coding conditions:** The results are presented for four target compression ratios (CRs): 10, 20, 40, and 100 [8]. To guarantee a fair subjective quality comparison with the anchors results provided for the JPEG [8], the target coding bits (corresponding to each target CR) is considered to be the maximum allowed amount of bits to encode the test images with the proposed LFC (Bi-SS) solution. This consideration also corresponds to the case where the target coding bits are used to define a restriction in the storage or channel capacity. Therefore, for each RD point, the quantization parameter (QP) [9] of the proposed encoder is adjusted to have the closest number of coding bits that is equal to or smaller than the target coding bits specified by the target CR.
- **Performance Evaluation:** The objective evaluation was performed using the procedure outlined in [8]. In this paper, the results are shown in terms of the mean YUV PSNR, $PSNR_{YUV_{mean}}$, and mean Y PSNR, $PSNR_{Y_{mean}}$, of all rendered viewpoint images [8]. Therefore, the Bjøntegaard Delta (BD) [15] results in terms of PSNR and rate (BR) are presented in Tables 1 and 2.

Table 1. BD performance in terms of $\boldsymbol{PSNR_{YUV_{mean}}}$(PSNR) and coding bits (BR) for the proposed *LFC (Bi-SS)* solution

Image ID	*JPEG*		*HEVC*		*LFC (Uni-SS)*	
	PSNR [dB]	**BR [%]**	**PSNR [dB]**	**BR [%]**	**PSNR [dB]**	**BR [%]**
I01	5.13	-66.57	1.06	-29.89	0.38	-12.70
I02	5.48	-70.56	0.68	-20.82	0.25	-8.60
I03	4.40	-58.06	0.23	-6.83	0.09	-2.91
I04	5.18	-62.25	0.25	-8.93	0.05	-1.76
I05	4.22	-71.75	0.69	-35.28	0.29	-19.53
I06	5.49	-79.12	1.52	-59.99	0.63	-42.55
I07	4.47	-68.64	0.37	-15.75	0.12	-5.79
I08	4.26	-72.60	0.73	-37.62	0.46	-29.04
I09	5.88	-78.74	1.54	-45.26	0.34	-13.50
I10	4.00	-72.66	0.15	-11.69	0.05	-5.33
I11	6.19	-85.78	1.74	-66.22	0.53	-34.10
I12	7.15	-81.36	1.61	-53.07	0.42	-22.19
Average	**5.15**	**-72.34**	**0.88**	**-32.61**	**0.30**	**-16.50**

Table 2. BD performance in terms of $\boldsymbol{PSNR_{Y_{mean}}}$(PSNR) and coding bits (BR) for the proposed *LFC (Bi-SS)* solution

Image ID	*JPEG*		*HEVC*		*LFC (Uni-SS)*	
	PSNR [dB]	**BR [%]**	**PSNR [dB]**	**BR [%]**	**PSNR [dB]**	**BR [%]**
I01	5.65	-65.67	1.21	-29.73	0.45	-13.08
I02	5.90	-69.22	0.79	-21.12	0.27	-8.06
I03	4.71	-55.18	0.25	-6.56	0.09	-2.49
I04	5.30	-59.83	0.32	-9.39	0.07	-2.10
I05	4.71	-70.43	0.76	-34.26	0.30	-18.21
I06	5.79	-77.41	1.71	-56.86	0.67	-39.14
I07	4.56	-68.32	0.37	-15.13	0.11	-4.97
I08	4.31	-69.83	0.80	-37.13	0.47	-27.78
I09	6.70	-77.11	1.82	-44.88	0.37	-12.66
I10	4.14	-71.12	0.17	-13.36	0.07	-7.23
I11	6.57	-85.17	1.88	-66.97	0.56	-34.17
I12	8.21	-80.46	1.80	-52.96	0.43	-20.89
Average	**5.55**	**-70.81**	**0.99**	**-32.36**	**0.32**	**-15.90**

The following three benchmark solutions are presented and compared with the proposed LFC (Bi-SS) solution:

1) *JPEG:* The reconstructed viewpoint images provided in [8] (anchors) were used for comparison.
2) *HEVC:* The test images were encoded with HEVC, using the Main Still Picture profile [9].
3) *LFC (Uni-SS):* The test images were encoded with the solution proposed in [13], where no bi-prediction is used for the SS estimation and compensation.

The reference software of HEVC version 14.0 was used as the benchmark, as well as the base software for implementing the proposed codec.

As can be seen by the results in Table 1, the proposed LFC (Bi-SS) solution presents significantly better RD performance compared to the JPEG (up to 7.15 dB, with 81.36 % of bit savings) and HEVC (up to 1.74 dB, with

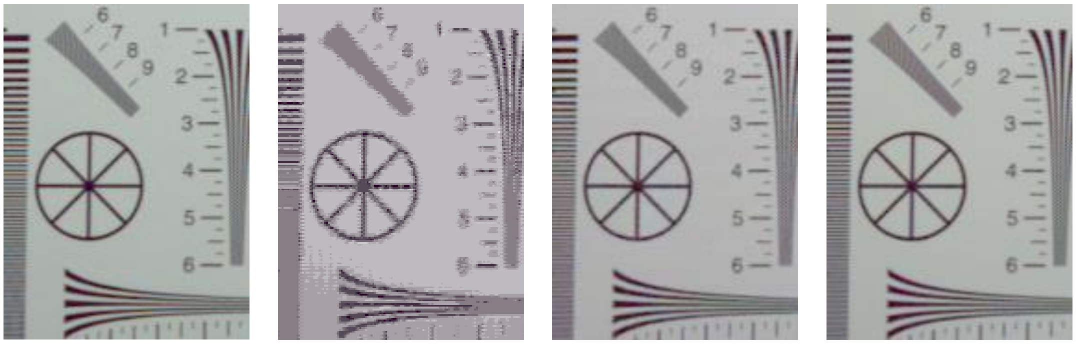

Fig. 3 Comparison of a portion from the central viewpoint image of ISO_Chart_12 [8] rendering (from left to right): original image; compressed image using JPEG; compressed image using HEVC; and compressed image using the proposed LFC (Bi-SS) solution

66.22 % of bit savings). Moreover, it is possible to conclude that introducing the proposed bi-predicted SS estimation and compensation leads to further improvements compared to LFC (Uni-SS) (up to 0.53 dB, with 34.10 % of bit savings).

Comparing the results from Tables 1 and 2, it is possible to see that the proposed LFC (Bi-SS) presents even more evident gains when considering only the luma component (for the $PSNR_{Y_{mean}}$ .metric). Furthermore, from a visual inspection in the viewpoint images rendered from the coded light field with the different coding solutions at the same bitrate (CR equal to 100), it is possible to conclude that the proposed LFC (Bi-SS) presents considerably better visual quality than JPEG and improvements are also noticeable when compared to HEVC. Fig. 3 illustrates this for a portion of the central viewpoint from test image I12 [8].

4. CONCLUSIONS

This paper proposed a light field image codec solution based on HEVC and using a bi-predicted self-similarity (SS) estimation and compensation. As discussed in this paper, the proposed light field image coding solution is shown to be advantageous in terms of the simplicity of the coding format, while keeping the encoder/decoder complexity and memory load comparable to HEVC. Moreover, the proposed bi-predicted SS estimation and compensation led to significantly superior performance compared to JPEG (up to 7.15 dB / -85.78 %), HEVC (up to 1.74 dB/ -66.22 %), and the previous SS-based solution (up to 0.56 dB / -34.17 %). Future work includes the study of the tradeoff between coding performance and the complexity by varying the search window and/or for generalizing to an N-predicted SS light field coding scheme.

5. ACKNOWLEDGMENT

The authors acknowledge the support of FCT, under UID/EEA/50008/2013 project and SFRH/BD/79480/2011 grant.

6. REFERENCES

[1] M. Levoy and P. Hanrahan, “Light Field Rendering,” in *SIGGRAPH 96*, 1996, pp. 31–42.

[2] A. Aggoun, E. Tsekleves, M. R. Swash, D. Zarpalas, A. Dimou, P. Daras, P. Nunes, and L. D. Soares, “Immersive 3D holoscopic video system,” *IEEE Multimed.*, vol. 20, no. 1, pp. 28–37, 2013.

[3] J. Arai, “Integral Three-Dimensional Television (FTV Seminar),” ISO/IEC JTC1/SC29/WG11 M34199, Sapporo, Japan, Jul. 2014.

[4] R. Raghavendra, K. B. Raja, and C. Busch, “Presentation attack detection for face recognition using light field camera.,” *IEEE Trans. Image Process.*, vol. 24, no. 3, pp. 1060–75, Mar. 2015.

[5] X. Xiao, B. Javidi, M. Martinez-Corral, and A. Stern, “Advances in Three-Dimensional Integral Imaging: Sensing, Display, and Applications [Invited],” *Appl. Opt.*, vol. 52, no. 4, pp. 546–560, Feb. 2013.

[6] P. Schelkens, “JPEG PLENO – Scope, use cases and requirements Ver.1.3,” ISO/IEC JTC 1/SC 29/WG1 M71003, La Jolla, CA, USA, 2016.

[7] M. P. Tehrani, S. Shimizu, G. Lafruit, T. Senoh, T. Fujii, A. Vetro, and M. Tanimoto, “Use Cases and Requirements on Free-viewpoint Television (FTV),” ISO/IEC JTC1/SC29/WG11 MPEG N14104, Geneva, Switzerland, Oct. 2013.

[8] M. Rerabek, T. Bruylants, T. Ebrahimi, F. Pereira, and P. Schelkens, “Call for Proposals and Evaluation Procedure.” ICME 2016 Grand Challenge: Light Field Image Compression, Seattle, USA, pp. 1–8, 2016.

[9] G. J. Sullivan, J. Ohm, W.-J. Han, T. Wiegand, and T. Wiegand, “Overview of the High Efficiency Video Coding (HEVC) Standard,” *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1649–1668, 2012.

[10] K. Ugur and J. Lainema, “Updated results on HEVC still picture coding performance,” JCTVC-M0041, Incheon, KR, 2013.

[11] P. Hanhart, M. Rerabek, P. Korshunov, and T. Ebrahimi, “Subjective evaluation of HEVC intra coding for still image compression,” JCTVC-L0380, Geneva, 2013.

[12] C. Conti, P. Nunes, and L. D. Soares, “New HEVC Prediction Modes for 3D Holoscopic Video Coding,” in *2012 19th IEEE International Conference on Image Processing*, 2012, pp. 1325–1328.

[13] C. Conti, L. D. Soares, and P. Nunes, “HEVC-based 3D holoscopic video coding using self-similarity compensated prediction,” *Signal Process. Image Commun.*, vol. 42, pp. 59–78, Jan. 2016.

[14] M. Flierl and B. Girod, *Video Coding with Superimposed Motion-Compensated Signals - Applications to H.264 and Beyond*, 1st ed. Springer US, 2004.

[15] G. Bjøntegaard, “Calculation of Average PSNR Differences between RD Curves,” VCEG-M33, Austin, TX, USA, Apr. 2001.